CLAIMS

WHAT IS CLAIMED IS:

1. A system for high throughput detection of genotypes comprising

a sample preparation automation system;

a sample tracking system;

an automated high density probe array loader;

a computer system for managing hybridization data and for analyzing hybridization data to make genotype calls.

2. The system of Claim 1 wherein the sample preparation automation system is a robotic device for handling multwell plates.

3. The system of Claim 1 wherein the sample tracking system is a bar code system.

4. The system of Claim 1 wherein computer system comprises a processor; and a memory being coupled with the processor, the memory storing a plurality of machine instructions that cause the processor to perform the method step of analyzing the hybridization to determine the genotype, wherein the analyzing comprises calling a genotype by calculating the likelihood of a set of models for the hybridization and the base is called based upon the likelihood of the models; wherein the distribution of hybridization intensities are assumed to be Gaussein and forward and reverse strand are treated as independent replicates.

5. The method of Claim 4 wherein the models are five homozygote Models (Null, A, C, G, T) for a haploid and and 11 models (Null, A, C, G, T, A-C, A-G, A-T, C-G,

C-T, G-T) for a diploid.

6. The method of Claim 5 wherein likelihood of a model is calculated independently for both the forward and reverse strands and is combined for the overall likelihood of the model.

7. The method of Claim 6 wherein a genotype is called if one model fits the hybridization data better than all other models.

8. A method for determining the genotype of a polymorphism comprising:

preparing a nucleic acid sample;

determining the hybridization of the nucleic acid sample with a high density oligonucleotide probe array; wherein the high density oligonucleotide probe array having probes interrogating the polymorphism; and

analyzing the hybridization to determine the genotype, wherein the analyzing comprises calling a genotype by calculating the likelihood of a set of models for the hybridization and the base is called based upon the likelihood of the models.

9. The method of Claim 8 wherein the models are five homozygote Models (Null, A, C, G, T) for a haploid and and 11 models (Null, A, C, G, T, A-C, A-G, A-T, C-G, C-T, G-T) for a diploid.

10. The method of Claim 9 wherein likelihood of a model is calculated independently for both the forward and reverse strands and is combined for the overall likelihood

of the model.

11. The method of Claim 10 wherein a genotype is called if one model fits the hybridization data better than all other models.

12. The method of Claim 11 wherein the likelihood of a set of hybridization intensity as measured by pixel intensities is: .

$$\ln(L) = -\frac{1}{2}\sum N_x[\ln(\hat{\sigma}_x^2) + (V_x + M_x^2 - 2\hat{\mu}_x M_x + \hat{\mu}_x^2)\ /\hat{\sigma}_x^2 + \ln(2\pi)].$$

wherein N_x is the number of pixels observed in feature x ; V_x is the observed variance for feature x, M_x is the observed mean for feature x, μ_x is the estimated mean for feature x under a model, and σ^2_x is the estimated variance for feature x, and wherein the sum is taken over all features x, where x is either A, C, G, or T, on the forward and reverse strands.

13. The method of Claim 12 wherein the mean and variance for a Null Model are estimated according to:

$$\hat{\mu}_r(b) = \frac{N_r(A)M_r(A) + N_r(C)M_r(C) + N_r(G)M_r(G) + N_r(T)M_r(T)}{N_r(A) + N_r(C) + N_r(G) + N_r(T)}$$

$$\hat{\mu}_f(b) = \frac{N_f(A)M_f(A) + N_f(C)M_f(C) + N_f(G)M_f(G) + N_f(T)M_f(T)}{N_f(A) + N_f(C) + N_f(G) + N_f(T)}$$

$$\hat{\sigma}_f^2(b) = \frac{\begin{array}{c}N_f(A)(V_f(A) + M_f^2(A)) + N_f(C)(V_f(C) + M_f^2(C)) \\ + N_f(G)(V_f(G) + M_f^2(G)) + N_f(T)(V_f(T) + M_f^2(T))\end{array}}{N_f(A) + N_f(C) + N_f(G) + N_f(T)} - \hat{\mu}_f^2(b)$$

$$\hat{\sigma}_r^2(b) = \frac{\begin{array}{c}N_r(A)(V_r(A) + M_r^2(A)) + N_r(C)(V_r(C) + M_r^2(C)) \\ + N_r(G)(V_r(G) + M_r^2(G)) + N_r(T)(V_r(T) + M_r^2(T))\end{array}}{N_r(A) + N_r(C) + N_r(G) + N_r(T)} - \hat{\mu}_r^2(b)$$

14. The method of Claim 10 wherein the mean and variance for a hymozygous model are estimated according to:

$$\widehat{\mu}_f(b) = \frac{N_f(C)M_f(C) + N_f(G)M_f(G) + N_f(T)M_f(T)}{N_f(C) + N_f(G) + N_f(T)}$$

$$\widehat{\sigma}_f^2(b) = \frac{N_f(C)\omega_f(C) + N_f(G)\omega_f(G) + N_f(T)\omega_f(T)}{N_f(C) + N_f(G) + N_f(T)}.$$

$$\omega_f(x) = V_f(x) + M_f^2(x) - 2M_f(x)\widehat{\mu}_f(b) + \widehat{\mu}_f(b) + \widehat{\mu}_f^2(b)$$

$$\widehat{\mu}_f(A) = M_f(A),$$

$$\widehat{\sigma}_f^2(A) = V_f(A).$$